The Annals of Applied Statistics
2011, Vol. 5, No. 1, 61–64
DOI: 10.1214/10-AOAS398G
Main article DOI: 10.1214/10-AOAS398

DISCUSSION OF: A STATISTICAL ANALYSIS OF MULTIPLE TEMPERATURE PROXIES: ARE RECONSTRUCTIONS OF SURFACE TEMPERATURES OVER THE LAST 1000 YEARS RELIABLE?

By Murali Haran and Nathan M. Urban

Pennsylvania State University and Princeton University

We thank the authors for a thought-provoking paper (henceforth MW). Their work may be divided into two parts: *reconstruction*, where the authors develop a Bayesian model for reconstructing historic temperatures based on proxies, along with associated measures of uncertainty; and *validation*, where they study how accurately their model corresponds to data by using cross-validation techniques or comparing proxies to simulated time series that are unrelated to temperature. We discuss both aspects of the paper although we focus mostly on reconstruction. While our comments may seem critical of MW, our views apply more generally to much of the existing work in this area.

We begin with a discussion of the reconstruction in MW. Given the advances in modeling for large, rich, complicated space–time processes and the availability of temperature proxies in the form of space–time data sets, we believe statistical approaches to paleoclimate reconstruction should make full use of such spatial data instead of using spatially aggregated forms of the data (as in MW). Such spatial aggregation may have the effects of removing interesting signals and of making it more difficult to define a credible error structure since proxy data are less directly related to global temperature than local temperature. This is an issue not only with MW, but also the reconstruction work of many others. In addition, recent advances in computationally efficient approaches for fitting hierarchical spatiotemporal models open up the possibility of developing more realistic models that account for various sources of error while incorporating specialized scientific knowledge into the models as appropriate [cf. Banerjee, Carlin and Gelfand (2004); Gelfand et al. (2010) and the references therein]. We believe that

Received September 2010.

such models are likely to provide more reliable estimates along with associated uncertainty estimates, both of which are important for drawing sound scientific conclusions.

We outline some ways in which we believe the model in MW can be improved upon.

(i) The authors approach this as a regression problem where they treat the proxies as the predictor and the temperature observations as response, and then use the proxies to extrapolate the temperature backwards. We believe it is more appropriate to view temperature as predicting proxies rather than the other way around. Recasting the problem in this "calibration" framework allows for more realistic models for measurement error and dependence. As is well known, ignoring measurement error in regressions can lead to erroneous conclusions [cf. Fuller (1980); Carroll, Ruppert and Stefanski (1995)].

(ii) The process by which MW selects proxies is problematic. We wonder why MW choose only those proxies that go all the way back given the availability of approaches for dealing with missing information.

(iii) The proxies are all very different in terms of scale, how they were collected and possibly aggregated, the kind of measurement error involved, and other characteristics such as spatial dependence. Also, the data associated with the proxies may be very different; for instance, some are discrete, some are continuous, and they may be at different frequencies. Critically, proxies are vastly different in terms of what they tell us about temperature; for example boreholes provide highly "smoothed" temperature reconstructions, while tree rings or lake varves can have annual resolution. It therefore seems inappropriate to merge all proxies together into a single regression model without accounting for their individual properties.

(iv) Principal Components Analysis (PCA) is a reasonable approach to reducing the dimensions of predictors in a regression problem, but we are concerned that PCA, like the LASSO, treats the paleo-reconstruction problem like a "data mining" problem, that is, a problem where nothing is known about underlying relationships among the predictors and the temperature field. For instance, negative regression coefficients may not be tenable in several cases.

(v) It may be possible to construct more realistic proxy-temperature relationships using process models [cf. Guiot et al. (2009)], although this may be more feasible for stronger climate signals, for example, deglaciations, than those present in the late Holocene.

We agree with the authors that if proxies show no recent changes then they may be inappropriate for extrapolating backwards. However, any discussion of Holocene paleoclimate reconstructions should be kept in perspective. It is tempting to use paleodata to make inferences about future climate. There have been attempts to use Holocene paleodata to constrain the climate sensitivity [cf. Hegerl et al. (2006); Schneider (2007) for discussions

of related methodology], though they do not account for temporal or spatial dependence, and share the limitations of the paleo-reconstructions upon which they are based. But independent of the accuracy of a paleotemperature reconstruction, there are limits to what past climates can tell us about possible future climates. Over the next few centuries the climate system will likely be strongly forced by continued greenhouse gas emissions. By contrast, the Holocene climate was relatively weakly forced, and not primarily by greenhouse gases. Given these differences, it is unclear to what extent further refinement of millennial temperature reconstructions can contribute to questions about the future climate. However, this does not detract from their potential usefulness in answering questions about natural climate variability, such as spatiotemporal patterns and attribution of past climate change.

Regarding the validation approach in MW: while we appreciate the principle behind comparing proxy-based reconstructions with constructions based on randomly generated proxies, it is perhaps not entirely surprising that models with dependent errors are good interpolators over short time periods. The actual proxies themselves may not be as good for short time periods, especially in the case of low-frequency proxies like a borehole, or an ecological proxy like a tree ring which might be confounded by subdecadal non-temperature variability [this is related to issue (iii) above]. One might also believe that a proxy would perform better at extrapolation over longer time periods.

In summary, we do not argue for or against the conclusions of this paper as much as we argue that much of the statistical work done in this paper and other related papers do not take full advantage of existing data, scientific knowledge and the latest in statistical methods, particularly hierarchical space–time modeling [see Tingley et al. (2010) for a discussion of possible strategies to pursue]. Having said that, the researchers in this field deserve much credit for their pioneering work on temperature reconstructions which has laid the foundations for an important and interesting field of research. We are delighted that more statisticians are becoming involved in the statistical aspects of climate science and we commend the authors for taking on this challenging problem in a methodical fashion. We particularly like their method of carefully working through both reconstruction and validation; this two-pronged approach provides a nice template for future work.

Acknowledgment. We thank Don Richards for helpful discussions.

REFERENCES

Banerjee, S., Carlin, B. and Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall, London.

Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models. Monographs on Statistics and Applied Probability* **63**. Chapman & Hall, London. MR1630517

Fuller, W. A. (1980). Properties of some estimators for the errors-in-variables model. *Ann. Statist.* **8** 407–422. MR0560737

Gelfand, A. E., Diggle, P. J., Fuentes, M. and Guttorp, P. (2010). *Handbook of Spatial Statistics*. CRC Press, Boca Raton, FL.

Guiot, J., Wu, H. B., Garreta, V., Hatté, C. and Magny, M. (2009). A few prospective ideas on climate reconstruction: From a statistical single proxy approach towards a multi-proxy and dynamical approach. *Climates of the Past Discussions* **5** 99–125.

Hegerl, G. C., Crowley, T. J., Hyde, W. T. and Frame, D. J. (2006). Climate sensitivity constrained by temperature reconstructions over the past seven centuries. *Nature* **440** 1029–1032.

Schneider, T. (2007). Climate modelling: Uncertainty in climate-sensitivity estimates. *Nature* **446** E1; discussion E2.

Tingley, M., Craigmile, P., Haran, M., Li, B., Mannshardt-Shamseldin, E. and Rajaratnam, B. (2010). Piecing together the past: Statistical insights into paleoclimatic reconstructions. Technical report, Dept. Statistics, Stanford Univ.

Department of Statistics
Pennsylvania State University
University Park, Pennsylvania 16802
USA
E-mail: mharan@stat.psu.edu

Woodrow Wilson School of Public
and International Affairs
Princeton University
Princeton, New Jersey 08544
USA
E-mail: nurban@princeton.edu